



```

RESULT 2
VCLA_GOSHI STANDARD; PRT; 605 AA.
AC P09799;
DT 01-MAR-1989 (rel. 10, Created)
DT 15-JUL-1999 (rel. 38, Last sequence update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
  Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
  Malvales; Malvaceae; Gossypium.
RN (1)
RP SEQUENCE FROM N.A.
RA Chlan C.A., Borroto K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
  germination. XIX. Sequences and genomic organization of the alpha
  globulin (vicilin) genes of cottonseed."
RL Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC MEMBRANE-BOUND VACUOLAR PROTEIN
  BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
  CONVULICLIN, CONGLICININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M19378; AAA33069.1; -.
DR PIR: S06398; S06398.
DR HSSP: P50477; 1CAK.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore_7s; 1.
KW Seed storage protein; Signal.
FT SIGNAL 1 23
FT CHAIN 24 605 VICILIN GC72-A.
FT SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;

Query Match
Best Local Similarity 32.0%; Score 120; DB 1; Length 605;
Matches 33; Conservative 14; Mismatches 16; Indels 42; Gaps 5;

QY 3 DPQQRREYEDCCRRHCEQOEPRLL-----QY-----QCQRQOEQORONG 39
DB 79 DPQRR-YQDCRQHCQOEERLRPHCEQSCREYEQOQOQDPKQKRECOQRCQMOEORPE 137

QY 40 R-----GCDLNPORQSGRYE-----EGEEKOSDNP 66
DB 138 RQQCVKCECREQYQEDPMKGERENKWRKEEERESDEGEOQQRNRP 182

RESULT 3
GLH1_CAEEL STANDARD; PRT; 763 AA.
AC P34689; Q9TXH4;
DT 01-FEB-1994 (rel. 28, Created)
DT 01-OCT-2000 (rel. 40, Last sequence update)
DT 01-OCT-2000 (rel. 40, Last annotation update)
DE ATP-DEPENDENT RNA HELICASE GLH-1 (GERMLINE HELICASE-1).
CN GLH-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
  Rhabditidae; Peloderinae; Caenorhabditis.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;

```

```

RX MEDLINE-94022363; PubMed-8415696;
RA Russell D.L., Bennett K.L.;
RT "glh-1, a germ-line putative RNA helicase from Caenorhabditis, has
  four zinc fingers."
RL Proc. Natl. Acad. Sci. U.S.A. 90:9300-9304(1993).
RN (2)
RP REVISIONS TO 83-138; 275; 288 AND 398.
RA Russell D.L., McCrone J.S., Smith P.A., Gruidl M.E., Bennett K.L.;
RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PROBABLE ATP-BINDING RNA HELICASE.
CC -1- DEVELOPMENTAL STAGE: DURING GERM-LINE PROLIFERATION.
CC -1- SIMILARITY: TO OTHER "DEAD" BOX FAMILY HELICASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
DR EMBL: I19948; AAC27384.1; -.
DR PIR: A48686; A48686.
DR HSSP: P05888; 1AAR.
DR INTERPRO: IPR000629; -.
DR INTERPRO: IPR001410; -.
DR INTERPRO: IPR001650; -.
DR INTERPRO: IPR001878; -.
DR PFAM: PF00270; DEAD; 1.
DR PFAM: PF00271; helicase_C; 1.
DR PFAM: PF00098; zf-CCHC; 4.
DR PRINTS: PR00939; C2HCZNFINGER.
KW Helicase; ATP-binding; RNA-binding; Zinc-finger; Repeat.
DR PROSITE: PS00039; DEAD_ATP_HELICASE; 1.
FT DOMAIN 24 93
FT REPEAT 24 33
FT REPEAT 34 43
FT REPEAT 44 53
FT REPEAT 54 63
FT REPEAT 64 73
FT REPEAT 74 83
FT REPEAT 84 93
FT DOMAIN 207 236 GLY-RICH.
FT ZN_FING 160 173 C2HC-TYPE.
FT ZN_FING 185 198 C2HC-TYPE.
FT ZN_FING 244 257 C2HC-TYPE.
FT ZN_FING 264 277 C2HC-TYPE.
FT NP_BIND 385 392 ATP (BY SIMILARITY).
FT SITE 489 502 DEAD BOX.
FT SEQUENCE 763 AA; 79792 MW; ADB69DE286A028D6 CRC64;

Query Match
Best Local Similarity 20.8%; Score 78; DB 1; Length 763;
Matches 22; Conservative 11; Mismatches 26; Indels 18; Gaps 3;

QY 2 RPDQREYEDCCRRHCEQOEPRLLQYQO-----RRCQEOQORH-----GRGDL 44
DB 164 QCPGHRH-SDCPREKRERPRVCYNQCPGHRHRECTEERKPREGRTGGFGGAGCGNNG 222

QY 45 MNPQRCGSGRYEGEEK 61
DB 223 GNDGFGDGDFGGGEER 239

RESULT 4
INVO_CEBAL STANDARD; PRT; 428 AA.
AC P24709;
DT 01-MAR-1992 (rel. 21, Created)
DT 01-MAR-1992 (rel. 21, Last sequence update)
DT 01-FEB-1996 (rel. 33, Last annotation update)
DE INVOLUCRIN.

```

GN	IVL.
OS	Cebus albifrons (White-fronted capuchin).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Cebus.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=OEESOPHAGEAL FIBROBLAST;
RX	MEDLINE=92114750; PubMed=1766560;
RA	Phillips M., Rice R.H., Djian P., Green H.;
RT	"The involucrin genes of the white-fronted capuchin and cotton-top
RL	tamarin: the plathyrrhine middle region.";
CC	Mol. Biol. Evol. 8:579-591(1991).
CC	-I- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS
CC	IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
CC	MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE
CC	FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC	-I- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC	OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; M67478; AAA35405.1; -.
DR	INTERPRO; IPR000354; -.
DR	INTERPRO; IPR002360; -.
DR	PFAM; PF00904; Invoculcrin; 28.
DR	PROSITE; PS00795; INVOLUCRIN; 1.
DR	Keratinocyte; Repeat.
QO	SEQUENCE 428 AA; 50126 MW; 03AB7A15FC34EB02 CRC64;

```

Query Match      20.5% Score 77; DB 1; Length 428;
Best Local Similarity 29.2%; Pref. No. 0.5%;
Matches 21; Conservative 17; Mismatches 18; Indels 16; Gaps 3;

QY      8 EYEDCRHCEQDEERLQY-----OCQRCEQQRQHQHGKGLMNP-QRGSGR 54
      : | : | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db      257 QQEDQKLHLEQDEQLEHLEQDEGEKLHLEQCEQGLEHLEQDEQ---LELPDQVGQSK 313
      : | : | | | : | : | : | : | : | : | : | : | : | : | : | : |

QY      55 YESEGEQSDNP 66
      : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db      314 HLEQEKQLHP 325

RESULT      5
INVO_SAGE
ID      INVO_SAGE      STANDARD;      PRT;      493 AA.
AC      P24712;
DT      01-MAR-1992 (Rel. 21, Created)
DT      01-MAR-1992 (Rel. 21, Last sequence update)
DT      01-FEB-1996 (Rel. 33, Last annotation update)
DE      INVOLUCRIN.
GN      IVL.
OS      Sagulnus oedipus (cotton-top tamarin).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Sagulnus.
[1]
RN      RP
RP      SEQUENCE FROM N.A.
RC      TISSUE=VAGINAL FIBROBLAST;
RX      MEDLINE=92114750; Pubmed=1766360;
RA      Phillips M., Rice R.H., Djian P., Green H.;
RT      "The involucrin genes of the white-fronted capuchin and cottontop
RT      tamarin: the platyrrhine middle region.";
RL      Mol. Biol. Evol. 8:579-591(1991).
CC      -I- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS
CC      IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
CC      MEMBRANE PROTEINS BY TRANSGUTAMINASE. ALL THAT RESULTS IN THE
CC      FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC

```

```

CC -1- TISSUESPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M67477; AAA36950.1; -.
DR HSSP; P80220; IDIP.
DR INTERPRO; IPR000354; -.
DR INTERPRO; IPR002360; -.
DR PFAM; PF00904; Involucrin;34.
DR PROSITE; PS00795; INVOLUCRIN; 1.
KW Keratinocyte. Repeat.
SQ SEQUENCE 493 AA; 57920 MW; A626FCAE7799AE09 CRC64;

```

	Query Match	20.3%;	Score 76;	DB 1;	Length 493;	
	Best Local Similarity	36.6%;	Pred. No. 0.75;			
	Matches	26; Conservative	11;	Mismatches	20;	Indels 14; Gaps 4;
OY	8 EYEDCRHCEQDEPRLYOOCRCO----	EEO-----	RHGKRGDLMNP-QRGSGRY	55		
	: :           :   :		:   :			
Dd	316 EOELQLKLHEOEDELKHLYDEGOLLEHLDEGOLKHLBQH--	BGOLELPEOOVGOSKH	373			
OY	56 EEGEKOSDNP	66				
	:   :					
Lb	374 LEOEKOENP	384				

RESULT	6
BLISA_HUMAN	
ID	BLISA_HUMAN
AC	002832;
DT	01-JUN-1994 (Rel. 29, Created)
DT	01-JUN-1994 (Rel. 29, Last sequence update)
DT	01-OCT-1994 (Rel. 30, Last annotation update)
DE	B-LYMPHOCYTE ANTIGEN PRECURSOR (B-LYMPHOCYTE SURFACE ANTIGEN) (721IP).
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=PLACENTA.
RX	MEDLINE=93066251; PubMed=1438229;
RA	Voland J.R., Wyzkowski R.J., Huang M., Dutton R.W.;
RT	Cloning and sequencing of a trophoblast-endothelial-activated
RT	lymphocyte surface protein: cdna sequence and genomic structure.*;
RL	Proc. Natl. Acad. Sci. U.S.A. 89:10425-10429(1992).
CC	- FUNCTION: THIS PROTEIN MAY BE INVOLVED IN B-CELL ACTIVATION.
CC	MAY ALSO BE INVOLVED IN SIGNAL TRANSDUCTION AND GENE REGULATION.
CC	- SUBUNIT: MONOMER.
CC	- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED (PROBABLE).
CC	- TISSUE SPECIFICITY: EXPRESSED IN ACTIVATED LYMPHOCYTES, MOST
CC	VASCULAR ENDOTHELIUM, AND SYNCTIORHOBIOBLAST.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
EMBL	M99578; AAA36187.1; -
PIR	PIR; A46419; A46419.
KW	Glycoprotein; Signal; Membrane.
FT	SIGNAL 1 21
CHAIN	22 550 B-LYMPHOCYTE ANTIGEN.

FT DOMAIN 344 365 LEUCINE-ZIPPER.  
 FT CAROHD 85 85 N-LINKED (GLCMAC. . .) (POTENTIAL).  
 FT CAROHD 138 138 N-LINKED (GLCMAC. . .) (POTENTIAL).  
 SQ SEQUENCE 550 AA: 64120 MW: 02989338A5FB80F7 CRC64:

Query Match  
 Best Local Similarity 20.1%; Score 75.5; DB 1; Length 550;  
 Matches 21; Conservative 7; Mismatches 16; Indels 17; Gaps 3;

OY 4 POOREYEDCRH-----CEOEPRLOYOCORCOEOROHGGRGDLMPORG 51  
 DB 470 PADRVYRLCEHHAPPPAPAPCRCPQGEPC-PPRCGRRSQKREKRGKG----PMGG 524

OY 52 S 52  
 DB 525 S 525

## RESULT 7

GLB1\_MAIZE STANDARD; PRT; 573 AA.

AC P15590;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GLOBULIN-1 S ALLELE PRECURSOR (GLB1-S) (7S-LIKE).  
 GN GLB1.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 CC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=CV. INBERD LINE VA26;  
 RA Belanger F.C., Kriz A.L.;  
 RT "Molecular characterization of the major maize embryo globulin encoded  
 by the Glb1 gene.";  
 RL Plant Physiol. 91:636-643(1989).  
 RN [2]  
 RP SEQUENCE OF 87-100.  
 RX MEDLINE=89374022; PubMed=2775172;

RA Kriz A.L.;  
 RT "Characterization of embryo globulins encoded by the maize Glb  
 genes.";  
 RL Biochem. Genet. 27:239-251(1989).  
 CC -I- PPM: THREE PROTEIN-PROCESSING STEPS OCCUR IN THE FORMATION OF THE  
 MAJURE PROTEIN FROM THE PRIMARY TRANSLATION PRODUCT.

CC -I- POLYMORPHISM: THE THREE MOST COMMONLY OCCURRING GLB1 ALLELES HAVE  
 THE DESIGNATION L, I, AND S FOR LARGE, INTERMEDIATE, AND SMALL  
 PROTEINS, RESPECTIVELY.  
 CC -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,  
 CONVICILIN, CONGLYCININ, ETC.).

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M24845; AAA33467.1; .  
 DR HSSP: P50477; ICAM.  
 DR MAIZEDB: 30181; .  
 DR INTERPRO: IPR00113; .  
 DR PFAM: PF00546; Seedstore\_7s; 1.  
 KW Seed storage protein; Signal.

FT SIGNAL 1 18 OR 21 (POTENTIAL).  
 FT PROPEP 19 86  
 FT CHAIN 87 573 GLOBULIN-1 S ALLELE.  
 FT CAROHD 349 349 N-LINKED (GLCMAC. . .) (POTENTIAL).  
 SQ SEQUENCE 573 AA: 65029 MW: 525EDID00A062976 CRC64;

Query Match.  
 Best Local Similarity 19.2%; Score 72; DB 1; Length 573;  
 Matches 23; Conservative 4; Mismatches 21; Indels 18; Gaps 4;

OY 12 CRHCE---QOEPRLOYOC--ORCOEOROHGGRGDLMPORGSG-----RYEG 58  
 DB 40 CYRRCDEPRWHQRPRLCEQREEREKROERSHREAD-----RSGEGSSEDERERECK 94  
 OY 59 EKKQSD 64  
 DB 95 EKKQSD 100

## RESULT 8

ANDR\_PANTR STANDARD; PRT; 911 AA.

AC O97775;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE ANDROGEN RECEPTOR (DIHYDROTESTOSTERONE RECEPTOR).  
 GN AR OR NR3C4.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98404153; PubMed=9732460;  
 RA Choong C.S., Kempainen J.A., Wilson E.M.;  
 RT "Evolution of the primate androgen receptor: a structural basis for  
 disease.";  
 RL J. Mol. Evol. 47:334-342(1998).  
 CC -I- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN  
 THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR  
 PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.  
 CC -I- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -I- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
 A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
 CC -I- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 CC NR3 SUBFAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U94177; AAC73048.1; .  
 DR HSSP: P06536; IRGD.  
 DR INTERPRO: IPR000536; .  
 DR INTERPRO: IPR001103; .  
 DR INTERPRO: IPR001628; .  
 DR PFAM: PF00104; hormone\_rec; 1.  
 DR PFAM: PF00105; zf-C4; 1.  
 DR PRINTS: PR00047; STEROIDFINGER.  
 DR PRINTS: PR00521; ANDROGENR.  
 DR PROSITE: PS00031; NUCLEAR\_RECEPTOR; 1.  
 KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
 CC Zinc-finger; Steroid-binding.

FT DOMAIN 1 549  
 FT DNA\_BIND 551 616 MODULATING (BY SIMILARITY).  
 FT ZN\_FING 551 571 C4-TYPE ZINC FINGERS (TWO).  
 FT ZN\_FING 587 611 C4-TYPE.  
 FT DOMAIN 682 911 LIGAND-BINDING.  
 FT DOMAIN 57 78 POLY-GLN.  
 FT DOMAIN 84 88 POLY-GLN.  
 FT DOMAIN 192 196 POLY-GLN.  
 FT DOMAIN 371 380 POLY-PRO.  
 FT DOMAIN 395 401 POLY-ALA.  
 FT DOMAIN 448 464 POLY-GLY.

```

SQ      SEQUENCE# 911 AA: 98402 MW: 601B9BD4E697DAA4 CRC64;

Query Match          19.2%; Score 72; DB 1; Length 911;
Best Local Similarity 30.6%; Pred. No. 3.3;
Matches 19; Conservative 12; Mismatches 29; Indels 2; Gaps 1

OY      5 QOREFEDCRHRCGEQDEPRLQVCCQRRCQEQHQNHGCGILMNDPQSGSRVEEGEKQSD 64
Db       57 11:::-:-:11:::-:11:::-:11:::-:11:::-:11:::-:11:::-:11:::-:11:::-:
           00000000-000000000000000000ETSPRQ0000GEGSGPSQAHRNRGPXYLVLEDEQPS 114

OY      65 NP 66
Db       115 GP 116

RESULT   9
TRHY_HUMAN STANDARD; PRT: 1898 AA.
ID TRHY_HUMAN 007283;
AC 007283;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRICHOHYALIN
CN THH OR TRHY OR THL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
[1]
RX RP SEQUENCE FROM N.A. MEDLINE=93280194; Pubmed=7685034;
RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
RA Steinhert P.M.;
RT "The structure of human trichohyalin. Potential multiple roles as a
RT functional EF-hand-like calcium-binding protein, a cornified cell
RT envelope precursor, and an intermediate filament-associated (cross-
RT linking) protein.";
RT J. Biol. Chem. 268:12164-12176(1993).
[2]
RN RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=93315897; Pubmed=7686553;
RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinhert P.M.;
RT "Trichohyalin: a structural protein of hair, tongue, nail, and
RT epidermis.";
RL J. Invest. Dermatol. 101:65S-71S(1993).
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -1- SUBUNIT: MONOMER (PROBABLE).
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDILLA, AND IN
CC THE FLIPLOFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC THE EPIDERMIS.
CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STANDARD
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -1- PTM: KNOWN SUBSTRATE OF TRANSGUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDIYLARGININE DEIMIDASE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100
CC FAMILY.
```

```
CC -I- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-stb.ch/announce/
CC or send an email to license@isb-stb.ch).
CC -----
DR PIR: I09190; AAA65582.1; -.
DR PIR: A45973; A45973.
DR HSSP: P02633; 1BOC.
DR MIM: 190370; -.
DR INTERPRO: IPR001751; -.
DR INTERPRO: IPR002048; -.
DR PFAM: PF01023; S_100; 1.
DR PFAM: PF00036; ehand; 1.
DR PROSITE: PS00018; EF_HAND; 1.
DR PROSITE: PS00303; S100_CABP; 1.
KW Repeat; Calcium-binding.
FT DOMAIN 1 91 5-100 LIKE.
FT CA_BIND 22 33 SITE I (LOW AFFINITY) (POTENTIAL).
FT CA_BIND 62 73 SITE II (HIGH AFFINITY) (POTENTIAL).
FT DOMAIN 314 390 R-R-E-Q-E-E-R-R-E-Q-Q-L.
FT REPEAT 314 326 1-1 (APPROXIMATE).
FT REPEAT 327 339 1-2 (APPROXIMATE).
FT REPEAT 340 351 1-3 (APPROXIMATE).
FT REPEAT 352 364 1-4.
FT REPEAT 365 377 1-5.
FT REPEAT 378 390 1-6.
FT DOMAIN 391 444 9 X 6 AA TANDEM REPEATS OF R-R-E-Q-Q-L.
FT REPEAT 391 396 2-1.
FT REPEAT 397 402 2-2.
FT REPEAT 403 408 2-3.
FT REPEAT 409 414 2-4.
FT REPEAT 415 420 2-5.
FT REPEAT 421 426 2-6.
FT REPEAT 427 432 2-7.
FT REPEAT 433 438 2-8.
FT REPEAT 439 444 2-9.
FT DOMAIN 444 702 9 X 28 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 923 1162 8 X 30 AA TANDEM REPEATS.
FT REPEAT 923 952 4-1.
FT REPEAT 953 982 4-2.
FT REPEAT 983 1012 4-3.
FT REPEAT 1013 1042 4-4.
FT REPEAT 1043 1072 4-5.
FT REPEAT 1073 1102 4-6.
FT REPEAT 1103 1132 4-7.
FT REPEAT 1133 1162 4-8.
FT DOMAIN 1250 1849 23 X 26 AA APPROXIMATE TANDEM REPEATS.
FT CONFLICT 1752 1752 F-> L (IN REF. 2).
FT CONFLICT 1794 1801 OERDQRYR-> RSFTGSGTG (IN REF. 2).
FT CONFLICT 1857 1857 O-> K (IN REF. 2).
FT CONFLICT 1880 1880 V-> G (IN REF. 2).
SQ SEQUENCE 1898 AA; 247219 MM; A74B5947FB62E51D CRC64;

Query Match 19.1%; Score 71.5; DB 1; Length 1898;
Best Local Similarity 27.8%; Pred. No. 7.2;
Matches 20; Conservative 17; Mismatches 24; Indels 11; Gaps 3;
```

```

RESULT 10
INVO_AOTTR STANDARD: PRT: 544 AA.
ID INVO_AOTTR
AC P24708:
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE INVOLUCRIN.
GN IVL.
OS Aotus trivirgatus (Night monkey) (Douroucoulli).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
RN [1]
RX MEDLINE=90014142; PubMed=2507864;
RA "The involucrin gene of the owl monkey: origin of the early region.";
RL Mol. Biol. Evol. 6:460-468(1989).
CC -I- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS
CC IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
CC MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE
CC FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC -I- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M25313; AAA35375.1; -
DR HSSP: P80320; IDIP.
DR INTERPRO: IPR000354; -
DR INTERPRO: IPR002360; -
DR PfAM: PF00904; INVOLUCRIN; 42.
DR PROSITE: PS00795; INVOLUCRIN; 1.
DR KeraLncyte; Repeat.
SQ SEQUENCE 544 AA; 63927 MW; 2A02ABA5E1499FD CRC64;
Query Match 18.9%; Score 71; DB 1; Length 544;
Best Local Similarity 25.6%; Pred. No. 2.6;
Matches 21; Conservative 17; Mismatches 24; Indels 20; Gaps 2;
OY 5 QQRKEDCDRHCEQDEPRQY-----OCORRCQDEQHGKRGDL----- 44
DB 354 QDEEGGVKHLDEQDEQLKHLDEQGPVKHLQLEKLEHLDEQGLKHLQREQL 413
OY 45 MNPORGSGRYEGEGEKSDNP 66
DB 414 LPEQVGOSKHLQFEKQLHR 435
RESULT 11
ANDR_HUMAN STANDARD: PRT: 919 AA.
ID ANDR_HUMAN
AC P10275:
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ANDROGEN RECEPTOR (DIHYDROTESTOSTERONE RECEPTOR).
GN AR OR NR3C4 OR DHTF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RX MEDLINE=89112208; PubMed=3216866;
RA Lubahn D.B., Joseph D.R., Sar M., Tan J., Higgs H.N., Larson R.E.,
RA French F.S., Wilson E.M.;

```

```

RT "the human androgen receptor: complementary deoxyribonucleic acid
RT cloning, sequence analysis and gene expression in prostate.";
RL Mol. Endocrinol. 2:1265-1275(1988).
RN [2]
RX SEQUENCE FROM N.A.
RX MEDLINE=90083302; PubMed=2594783;
RA Lubahn D.B., Brown T.R., Simental J.A., Higgs H.N., Migeon C.J.,
RA Wilson E.M., French F.S.;
RT "Sequence of the intron/exon junctions of the coding region of the
RT human androgen receptor gene and identification of a point mutation
RT in a family with complete androgen insensitivity.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9534-9538(1989).
RN [3]
RX SEQUENCE FROM N.A.
RX MEDLINE=90258935; PubMed=2342476;
RA Govindan M.V.;
RT "Specific region in hormone binding domain is essential for hormone
RT binding and trans-activation by human androgen receptor.";
RL Mol. Endocrinol. 4:417-427(1990).
RN [4]
RX SEQUENCE FROM N.A.
RX TISSUE=PROSTATE;
RX MEDLINE=89017168; PubMed=3174628;
RA Chang C., Kokontis J., Liao S.;
RT "Structural analysis of complementary DNA and amino acid sequences of
RT human and rat androgen receptors.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:7211-7215(1988).
RN [5]
RX SEQUENCE FROM N.A.
RX TISSUE=PROSTATE;
RX MEDLINE=89098909; PubMed=2911578;
RA Tilley W.D., Marcelli M., Wilson J.D., McPhaul M.J.;
RT "Characterization and expression of a cDNA encoding the human androgen
RT receptor.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:327-331(1989).
RN [6]
RX SEQUENCE FROM N.A.
RX TISSUE=PROSTATE;
RX MEDLINE=91155943; PubMed=2293020;
RA Marcelli M., Tilley W.D., Wilson C.M., Griffin J.E., Wilson J.D.,
RA McPhaul M.J.;
RT "Definition of the human androgen receptor gene structure permits the
RT identification of mutations that cause androgen resistance: premature
RT termination of the receptor protein at amino acid residue 588 causes
RT complete androgen resistance.";
RL Mol. Endocrinol. 4:1105-1116(1990).
RN [7]
RX SEQUENCE OF 189-919 FROM N.A.
RX MEDLINE=88178111; PubMed=3353726;
RA Chang C., Kokontis J., Liao S.;
RT "Molecular cloning of human and rat complementary DNA encoding
RT androgen receptors.";
RL Science 240:324-326(1988).
RN [8]
RX SEQUENCE OF 468-919 FROM N.A.
RX MEDLINE=88240407; PubMed=3377788;
RA Trapman J., Klaassen P., Kuiper G.G.J.M., van der Korput J.A.G.M.,
RA Faber P.W., van Rooij H.C.J., Geurts van Kessel A., Voorhorst M.M.,
RA Mulder E., Brinkmann A.O.;
RT "Cloning, structure and expression of a cDNA encoding the human
RT androgen receptor.";
RL Biochem. Biophys. Res. Commun. 153:241-248(1988).
RN [9]
RX POLYMORPHISM OF POLY-GLN REGION.
RX MEDLINE=92220629; PubMed=1561105;
RA Sledzews H.F., Oostra B.A., Brinkmann A.O., Trapman J.;
RT "Tritnucleotide repeat polymorphism in the androgen receptor gene
RT (AR).";
RL Nucleic Acids Res. 20:1427-1427(1992).
RN [10]
RX POLYMORPHISM OF POLY-GLY REGION.
RX TISSUE=BLOOD;
RA Lu J., Danielsen M.;

```

RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
 RN [11]  
 RP VARIANTS SBWA IN POLY-GLN REGION.  
 RX MEDLINE=91287825; PubMed=2062380;  
 RA la Spada A.R., Wilson E.M., Lubahn D.B., Harding A.E., Fischback K.H.;  
 RT "Androgen receptor gene mutations in X-linked spinal and bulbar  
 RT muscular atrophy."  
 RL Nature 352:77-79(1991).  
 RN [12]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=95023089; PubMed=7937057;  
 RA Patterson M.N., Hughes L.A., Gottlieb B., Pinsky L.;  
 RT "The androgen receptor gene mutations database."  
 RL Nucleic Acids Res. 22:3560-3562(1994).  
 RN [13]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97169385; PubMed=9016528;  
 RA Gottlieb B., Trifiro M., Lumbroso R., Vasilion D.M., Pinsky L.;  
 RT "The androgen receptor gene mutations database."  
 RL Nucleic Acids Res. 25:158-162(1997).  
 RN [14]  
 RP VARIANT LNCAP ALA-877.  
 RX MEDLINE=91083633; PubMed=2260966;  
 RA Veldechoite J., Ris-Stalpers C., Kuiper G.G., Jenster G.,  
 RA Berrevoets C., Claassen E., van Rooij H.C.J., Trapman J.,  
 RA Brinkmann A.O., Mulder E.;  
 RT "A mutation in the ligand binding domain of the androgen receptor of  
 RT human LNCap cells affects steroid binding characteristics and  
 RT response to anti-androgens."  
 RL Biochem. Biophys. Res. Commun. 173:534-540(1990).  
 RN [15]  
 RP VARIANT CAIS MET-866.  
 RX MEDLINE=91186983; PubMed=2082179;  
 RA Brown T.R., Lubahn D.B., Wilson E.M., French F.S., Migeon C.J.,  
 RA Corfen J.L.;  
 RT "Functional characterization of naturally occurring mutant androgen  
 RT receptors from subjects with complete androgen insensitivity."  
 RL Mol. Endocrinol. 4:1759-1772(1990).  
 RN [16]  
 RP VARIANT CYS-774.  
 RX MEDLINE=91310758; PubMed=1856263;  
 RA Marcello M., Tilley W.D., Zoppi S., Griffin J.E., Wilson J.D.,  
 RA McPhaul M.;  
 RT "Androgen resistance associated with a mutation of the androgen  
 RT receptor at amino acid 772 (Arg-->Cys) results from a combination of  
 RT decreased messenger ribonucleic acid levels and impairment of  
 RT receptor function."  
 RL J. Clin. Endocrinol. Metab. 73:318-325(1991).  
 RN [17]  
 RP VARIANT CAIS ASN-695 AND HIS-695, AND SEQUENCE OF 629-723 FROM N.A.  
 RX MEDLINE=92131007; PubMed=1775137;  
 RA Ris-Stalpers C., Trifiro M.A., Kuiper G.G., Jenster G., Romalo G.,  
 RA Sai T., van Rooij H.C., Kaufman M., Rosenfield R.L., Liao S.;  
 RT "Substitution of aspartic acid-686 by histidine or asparagine in the  
 RT human androgen receptor leads to a functionally inactive protein with  
 RT altered hormone-binding characteristics."  
 RL Mol. Endocrinol. 5:1562-1569(1991).  
 RN [18]  
 RP VARIANTS CAIS AND PATS.  
 RX MEDLINE=93338440; PubMed=1307250;  
 RA Batch J.A., Williams D.M., Davies H.R., Brown B.D., Evans B.A.J.,  
 RA Hughes I.A., Patterson M.N.;  
 RT "Androgen receptor gene mutations identified by SSCP in fourteen  
 RT subjects with androgen insensitivity syndrome."  
 RL Hum. Mol. Genet. 1:497-503(1992).  
 RN [19]  
 RP VARIANT CAIS VAL-787.  
 RX MEDLINE=92235226; PubMed=1569163;  
 RA Nakao R., Haji M., Yanase T., Ogo A., Takayanagi R., Katsube T.,  
 RA Fukumaki Y., Nawata H.;  
 RT "A single amino acid substitution (Met-786-->Val) in the steroid-  
 RT binding domain of human androgen receptor leads to complete androgen  
 RT insensitivity syndrome."

RL J. Clin. Endocrinol. Metab. 74:1152-1157(1992).  
 RN [20]  
 RP VARIANT LNCAP ALA-877.  
 RX MEDLINE=92222955; PubMed=1562539;  
 RA Veldechoite J., Berrevoets C.A., Ris-Stalpers C., Kuiper G.G.,  
 RA Jenster G., Trapman J., Brinkmann A.O., Mulder E.;  
 RT "The androgen receptor in LNCap cells contains a mutation in the  
 RT ligand binding domain which affects steroid binding characteristics  
 RT and response to antiandrogens."  
 RL J. Steroid Biochem. Mol. Biol. 41:665-669(1992).  
 RN [21]  
 RP VARIANT MET-730.  
 RX MEDLINE=92335289; PubMed=1631125;  
 RA Newmark J.R., Hardy D.O., Tonb D.C., Carter B.S., Epstein J.L.,  
 RA Isaacs W.B., Brown T.R., Barrack E.R.;  
 RT "Androgen receptor gene mutations in human prostate cancer."  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:6319-6323(1992).  
 RN [22]  
 RP VARIANT CAIS VAL-754.  
 RX MEDLINE=93372806; PubMed=8103398;  
 RA Iobaccaro J.-M., Lumbroso S., Ktari R., Dumas R., Sultan C.;  
 RT "An exonic point mutation creates a Maellit site in the androgen  
 RT receptor gene of a family with complete androgen insensitivity  
 RT syndrome."  
 RL Hum. Mol. Genet. 2:1041-1043(1993).  
 RN [23]  
 RP VARIANT CAIS ARG-807.  
 RX MEDLINE=94108430; PubMed=8281140;  
 RA Adegemo O., Kallilo P.J., Palvimo J.J., Kontula K., Jaenke O.A.;  
 RT "A single-base substitution in exon 6 of the androgen receptor gene  
 RT causing complete androgen insensitivity: the mutated receptor fails  
 RT to transactivate but binds to DNA in vitro."  
 RL Hum. Mol. Genet. 2:1809-1812(1993).  
 RN [24]  
 RP VARIANT PATS VAL-743.  
 RX MEDLINE=93315568; PubMed=8325932;  
 RA Nakao R., Yanase T., Sakai Y., Haji M., Nawata H.;  
 RT "A single amino acid substitution (Gly743 --> Val) in the steroid-  
 RT binding domain of the human androgen receptor leads to Reifenstein

Query Match 18.9%; Score 71; DB 1; Length 919;  
 Best Local Similarity 29.0%; Pred. No. 4.2;  
 Matches 18; Conservative 14; Mismatches 28; Indels 2; Gaps 1;  
 QY 5 QOREVEDCCRHEDEQPRRQYOCORRCQROHGGGGLMPOROGSGRYEEGKOSD 64  
 DB 58 QQQ 115  
 QY 65 NP 66  
 DB 116 QP 117

RESULT 12  
 TAGB\_DICDI STANDARD; PRT; 1905 AA.  
 ID TAGB\_DICDI  
 AC P54683;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE PRESTALK-SPECIFIC PROTEIN TAGB PRECURSOR (EC 3.4.21.-).  
 GN TAGB.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Dictyostelida; Dictyostelium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX4;  
 RX MEDLINE=95262903; PubMed=7744252;  
 RA Shaulsky G., Kuspa A., Loomis W.F.;  
 RT "A multidrug resistance transporter/serine protease gene is required  
 RT for prestalk specialization in Dictyostelium."  
 RL Genes Dev. 9:1111-1122(1995).  
 CC -i- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGB MAY MEDiate

CC	INTRODUCTION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS.
CC	-1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE ATP-BINDING
CC	SB; ALSO KNOWN AS THE SUBTILASE FAMILY.
CC	-1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING
CC	TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.
CC	-1- SIMILARITY: STRONG, TO TAGC.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed, usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL: U20432; AAA62212.1; -.
DR	HSSP: P13569; INBD.
DR	DICTYDB: DD02059; TAGB.
DR	INTERPRO: IPR000209; -.
DR	INTERPRO: IPR001140; -.
DR	INTERPRO: IPR001617; -.
DR	PFAM: PF00664; ABC_membrane; 1.
DR	PFAM: PF00005; ABC_tran; 1.
DR	PRINTS: PR00723; SUBTILISIN.
DR	PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.
DR	PROSITE: PS00137; SUBTILASE_HIS; 1.
DR	PROSITE: PS00138; SUBTILASE_SER; 1.
DR	PROSITE: PS00211; ABC_TRANSPORTER; 1.
KM	Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;
KW	Signal.
FT	SIGNAL 1 ?
FT	CHAIN ? 1905
FT	DOMAIN 378 700
FT	DOMAIN ? 1905
FT	TRANSMEM 1011 1031
FT	TRANSMEM 1076 1096
FT	TRANSMEM 1121 1141
FT	TRANSMEM 1210 1230
FT	TRANSMEM 1309 1329
FT	TRANSMEM 1332 1352
FT	ACT_SITE 387 387
FT	ACT_SITE 432 432
FT	ACT_SITE 695 695
FT	NP_BIND 1553 1560
FT	DOMAIN 63 67
FT	DOMAIN 95 104
FT	DOMAIN 107 134
FT	DOMAIN 311 321
FT	DOMAIN 833 839
FT	DOMAIN 838 844
FT	DOMAIN 871 876
FT	DOMAIN 1012 1015
FT	DOMAIN 1386 1389
FT	DOMAIN 1398 1404
FT	DOMAIN 1445 1450
FT	DOMAIN 1765 1779
FT	DOMAIN 1782 1785
FT	DOMAIN 1807 1812
FT	DOMAIN 1815 1860
FT	DOMAIN 1872 1878
FT	CARBOHYD 594 594
FT	CARBOHYD 621 621
FT	CARBOHYD 742 742
FT	CARBOHYD 747 747
FT	CARBOHYD 823 823
FT	CARBOHYD 1172 1172
FT	CARBOHYD 1522 1522
FT	CARBOHYD 1658 1658
QO	SEQUENCE 1905 AA; 212518 MW; B8E223FA8B9AE13C CRC64;

[illegible]

RESULT	13			
AC	GLT_DROME	STANDARD:	PRT:	1023 AA.
ID	GLT_DROME			
AC	P33438;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	GLUTACTIN PRECURSOR.			
CN	GLT.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Eurygata; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	STRAIN-OREGON-R.			
RX	MEDLINE=90214632; PubMed=2108664;			
RA	Olson P.F., Fessler L.I., Nelson R.E., Sterne R.E., Campbell A.G.,			
RA	Fessler J.H.;			
RT	"Glutactin, a novel Drosophila basement membrane-related glycoprotein			
RT	with sequence similarity to serine esterases.";			
RL	EMBO J. 9:1219-1227(1990).			
CC	-1- FUNCTION: NOT KNOWN. BINDS CALCIUM IONS.			
CC	-1- SUBCELLULAR LOCATION: BASMENT MEMBRANES.			
CC	-1- PTR: EXTENSIVELY O-GLYCOSYLATED AND ALSO N-GLYCOSYLATED.			
CC	-1- PTR: FOUR TYROSINES ARE SULFATED.			
CC	-1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE TYPE-B			
CC	CARBOXYL-ESTERASE/LIPASE FAMILY.			
CC	-----			
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION			
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -			
CC	THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS			
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY			
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL			
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (SEE <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	OR SEND AN EMAIL TO <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL: X53286; CNA37380.1; -.			
DR	HSSP: P21836; 1MAH.			
DR	FLYBASE: FBgn0001114; GLT.			
DR	INTERPRO: IPR002018; -.			
DR	PFAM: PF00135; Coesterase; 2.			
DR	PROSITE: PS00941; CARBOXYL-ESTERASE-B_2; 1.			
KM	GLYCOPROTEIN; Sulfatation; Calcium-binding; Signal.			
FT	SIGNAL	1	17	
FT	CHAIN	18	1023	GLUTACTIN.
FT	STIMLAR	18	602	WITH FIVE-B CARBOXYL-ESTERASE/LIPASES.
FT	DOMAIN	603	615	POLY-THR.
FT	CARBOHD	115	115	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHD	368	368	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHD	402	402	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHD	810	810	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DISULFD	123	145	BY SIMILARITY.
FT	DISULFD	298	316	BY SIMILARITY.
FT	SEQUENCE	1023 AA;	118412 MW;	3638CFT9AB860E8C CRC64;

Query Match	18.9%;	Score 71;	DB 1;	Length 1905;	Query Match	18.7%;	Score 70;	DB 1;	Length 1023;
Best Local Similarity	25.4%;	Pred. No. 8.2;			Best Local Similarity	27.3%;	Pred. No. 5.8;		
					Matches .18;	Conservative 15;	Mismatches 33;	Indels 0;	Gaps 0.



OY 1 KRDPQREYEDCRHRCQDEPRLOVOCQRCQEQORQHGGDLMPORGSGRYEGEE 60  
 Db 716 ERTRQOQREYERQEQERQERERRELEERQERQEQOPEQPEYNEPMPGVPYQEP 775  
 OY 61 KQSDNP 66  
 Db 776 QPDDNP 781

## RESULT 14

NIT4\_NEUCR STANDARD; PRT; 1090 AA.  
 AC P28349;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE NITROGEN ASSIMILATION TRANSCRIPTION FACTOR NIT-4.  
 GN NIT-4.  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92017855; PubMed=1840634;  
 RA Yuan G.-F., Fu Y.-H., Marzluf G.A.;  
 RT "nit-4, a pathway-specific regulatory gene of Neurospora crassa,  
 RT encodes a protein with a putative binuclear zinc DNA-binding  
 RT domain."  
 RL Mol. Cell. Biol. 11:5735-5745(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92149315; PubMed=1531376;  
 RA Yuan G.-F., Marzluf G.A.;  
 RT "Molecular characterization of mutations of nit-4, the  
 RT pathway-specific regulatory gene which controls nitrate assimilation  
 RT in Neurospora crassa."  
 RL Mol. Microbiol. 6:67-73(1992).  
 CC -1- FUNCTION: PATHWAY-SPECIFIC REGULATORY GENE OF NITRATE  
 CC ASSIMILATION; IT ACTIVATES THE TRANSCRIPTION OF THE GENES FOR  
 CC NITRATE AND NITRITE REDUCTASES.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DOMAIN: THE GLUTAMINE-RICH DOMAIN MIGHT FUNCTION IN ACTIVATING  
 CC GENE EXPRESSION.  
 CC -1- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR  
 CC CLUSTER DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M80368; AAA3602.1; -  
 DR PIR: A41696; A41696.  
 DR PIR: S20033; S20033.  
 DR HSSP: P07272; 1PYI.  
 DR INTERPRO: IPR001138; -  
 DR PFAM: PF00172; Zn\_clns; 1.  
 DR PROSITE: PS00463; ZN2\_CY6\_FUNGAL\_1; 1.  
 DR PROSITE: PS50048; ZN2\_CY6\_FUNGAL\_2; 1.  
 KW Transcription regulation; Activator; DNA-binding; Nuclear protein;  
 KW Zinc; Metal-binding; Nitrate assimilation.  
 FT DNA\_BIND 53 81  
 FT DOMAIN 121 139 ZN(2)-CYS(6), FUNGAL-TYPE.  
 FT 213 229 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 429 450 ASP/GLU-RICH (ACIDIC).  
 FT 672 754 PRO-RICH.  
 FT DOMAIN 755 859 GLN-RICH.  
 FT 992 1024 POLY-GLN.  
 FT 98 98 K -> KP (IN REF. 1).  
 FT CONFLICT 467 467 L -> S (IN REF. 1).  
 SQ SEQUENCE 1090 AA; 120244 MW; 861DB9172EDD6114 CRC64;

Query Match 18.7%; Score 70; DB 1; Length 1090;  
 Best Local Similarity 30.8%; Pred. No. 6.2;  
 Matches 16; Conservative 14; Mismatches 18; Indels 4; Gaps 1;

## RESULT 15

LP61\_EIMTE STANDARD; PRT; 255 AA.  
 ID LP61\_EIMTE  
 AC P15714;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE ANTIGEN LPMC-61 (FRAGMENT).  
 OS Eimeria tenella.  
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriidae; Eimeriidae;  
 OC Eimeria.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SPOROZOITE;  
 RX MEDLINE=90348718; PubMed=2200963;  
 RA Ko C., Smith C.K. II, McDonnell M.;  
 RT "Identification and characterization of a target antigen of a  
 RT monoclonal antibody directed against Eimeria tenella merozoites."  
 RL Mol. Biochem. Parasitol. 41:53-64(1990).  
 CC -1- FUNCTION: UNKNOWN. THE GLN-RICH TANDM REPEATS MAY BE IMPORTANT  
 CC FOR AN UNKNOWN ASPECT OF THE PARASITIC LIFE CYCLE. MAY BE AN  
 CC IMPORTANT IMMUNOGEN.  
 CC -1- SUBUNIT: MAY BE COVALENTLY LINKED BY DISULFIDE BONDS TO OTHER  
 CC POLYPEPTIDES TO FORM THE 80 KDA ANTIGEN.  
 CC -1- DEVELOPMENTAL STAGE: PRESENT IN ALL STAGES THROUGHOUT THE  
 CC SPOULATION OF THE OOCYSTS AND IN THE SPOOROZOITES FOLLOWING  
 CC EXCYSTATION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M30933; AAA29079.1; -  
 DR PIR: A60637; A60637.  
 KW Antigen; Sporozoite; Repeat; Sporulation.  
 FT NON\_TER 1 1  
 FT DOMAIN 18 210 12 X APPROXIMATE TANDM REPEATS, GLN-  
 FT 18 210 RICH.  
 FT REPEAT 18 48  
 FT REPEAT 49 57 1.  
 FT REPEAT 58 65 2.  
 FT REPEAT 66 78 3.  
 FT REPEAT 79 90 4.  
 FT REPEAT 91 103 5.  
 FT REPEAT 104 140 6.  
 FT REPEAT 141 152 7.  
 FT REPEAT 153 164 8.  
 FT REPEAT 165 172 9.  
 FT REPEAT 173 192 10.  
 FT REPEAT 193 210 11.  
 FT NON\_TER 255 255  
 SQ SEQUENCE 255 AA; 31267 MW; 8C5E6005FFFC2DB3 CRC64;

Query Match 18.5%; Score 69.5; DB 1; Length 255;  
 Best Local Similarity 37.0%; Pred. No. 1.8;  
 Matches 17; Conservative 9; Mismatches 19; Indels 1; Gaps 1;

QY 4 POOREYEDCRRHCBOEPRLQYOCORRCOPOPOHOGRGGLMNPOR 49  
111: : 111:111: : 111: 111:  
Db 80 POOQPOMO-OEOWPQOQPOVQOQOQWPQOQHRRQHGQOQOQCMNSQO 124

Search completed: March 1, 2001, 16:16:50  
Job time: 430 Sec